The Value of Information from Sell-side Analysts

Linying Lv
Washington University in St. Louis

David Dreman

"Wall Street analysts know their companies. You should cut a research report in two. The first part, **the information about the company and its prospects**, is probably pretty good. The second part, **the recommendation**, should be used as kindling. We use analyst information, but we don't use the recommendations very often."

Structured Representation of Report Text

- I convert **unstructured analyst reports** into **structured embeddings**.

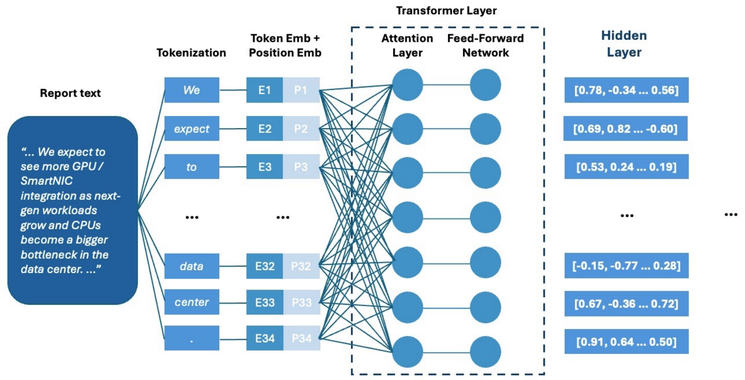


Fig 1. LLM embeddings architecture.

Text contains more information content than numbers.

- Train **Ridge models** each year and report **OOS R-squared** with distinct input

$$CAR_{[-1,+1],it} = \beta_0 + \beta' y^{\text{AI}}_{ijt} + \epsilon_{ijt},$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \left\| CAR_{[-1,+1],it} - \beta_0 - y^{\text{AI}}_{ijt}\beta \right\|_2^2 + \theta\|\beta\|_2^2 \right\}.$$

Year	**Rev only**	t-stat	**Text only**	t-stat	**Rev + text**	t-stat	t-stat	t-stat	t-stat
	(1)	(2)	(3)	(4)	(5)	(6)	(3)-(1)	(5)-(1)	(5)-(3)
2015	10.31%	4.27	12.63%	6.39	10.63%	2.98	3.63	0.16	-1.25
2016	14.61%	11.26	11.98%	3.93	17.08%	9.67	-1.15	3.82	2.93
2017	8.99%	6.23	11.11%	6.40	11.98%	7.09	4.99	5.96	2.38
2018	10.05%	3.28	10.87%	5.85	13.64%	5.95	0.87	5.77	4.47
2019	9.94%	20.14	12.16%	17.68	14.44%	26.17	2.68	19.77	4.83
2020	5.52%	3.88	3.82%	5.18	6.34%	6.16	-1.75	1.57	7.11
2021	5.43%	5.83	8.50%	6.21	11.94%	27.16	1.48	7.48	3.28
2022	9.78%	6.03	14.88%	10.34	16.95%	10.09	9.21	8.01	5.30
2023	6.68%	5.48	9.30%	4.17	10.76%	6.09	2.13	4.87	3.08
Overall	9.01%	9.45	10.19%	8.20	12.28%	8.87	1.66	3.95	3.77

Shapley Value Decomposition

- **Text Embeddings:**

$$y^{emb} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{e}_i,$$

- **Topic-specific Embeddings:**

$$y^{emb} = \sum_{p=1}^{P} y_p^{emb} = \frac{1}{N}\sum_{p=1}^{P}\sum_{i_p=1}^{N_p} e_{i_p},$$

- **Sentence-Segmented Text Embeddings:**

$$y^{emb} = \sum_{i=1}^{n} \frac{Token_i}{\sum_{i=1}^{n} Token_i} y_i^{emb},$$

- **Shapley Value of Topic p:**

$$\varphi_p\left(R^2_{\text{oos}}\right) = \sum_{S \subseteq P \setminus \{p\}} \frac{|S|!\,(P-|S|-1)!}{P!} \left[R^2_{\text{oos}}\left(y_s^{\text{emb}} + y_p^{\text{emb}}\right) - R^2_{\text{oos}}\left(y_s^{\text{emb}}\right)\right].$$

Income Statement Analyses is the most Important.

- The most valuable topics are **Income Statement Analyses** and **Financial Ratio Analyses**, each comprising 67% and 45% of text-based OOS R-squared.

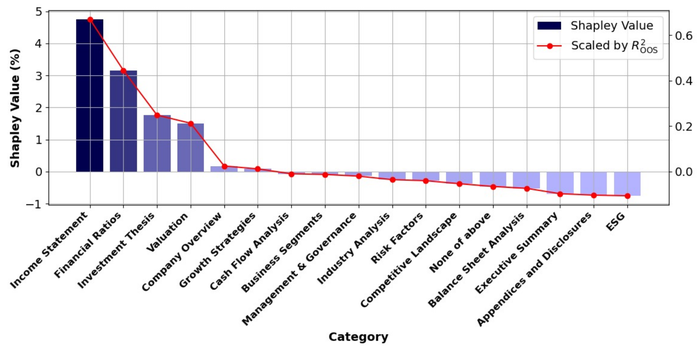


Fig 2. Topic importance

- The most valuable content is **analysts' interpretation of realized income**.
  - **Information Type**
    - Information Acquisition: Sentences that report quantitative financial data
      e.g. 'Reported EPS of $1.40 beat consensus of $0.94 and our estimate of $1.17.'
    - Information Interpretation: Sentences that analyze or interpret financial data, such as trends or market impacts.
      e.g. 'Biogen reported 1Q12 MS franchise sales below our estimate and the Street, as Avonex sales were negatively impacted by unfavorable distribution channel dynamics.'
  - **Time Reference**
    - Realized Income: Sentences that refer to historical results
      e.g. 'The company's revenue for the fiscal year 2022 was $500 million.'
    - Expected Income: Sentences that express future predictions
      e.g. 'The company expects to achieve revenue of $550 million in fiscal year 2023 based on current market conditions.'

Analyst information is profitable.

Analyst Information Value Estimation

The dollar value is the ratio of the explainable return variance—which quantifies the reduction in uncertainty attributable to analyst insights—to the price impact, representing the cost associated with trading on this information:

$$\widehat{\Omega}_{it} = \frac{r_{it}^2 - \left(r_{it} - \frac{\sum_{j=1}^{N}\widehat{r}_{ijt}}{N}\right)^2}{\widehat{\lambda}_{it}/p_{it\_}}.$$

- A stock typically has an average of **15 days** with report releases each year.
- The lower bound investors would expect to profit annually for early access to analyst information on an S&P 100 stock is **$6.89 million**.

	Mean	SE	95% CI	99% CI
Information Value ($M)	**0.47**	0.05	[0.38, 0.56]	[0.35, 0.58]
Information Value of Text ($M)	**0.38**	0.04	[0.30, 0.46]	[0.28, 0.48]
Information Value of Revisions ($M)	**0.34**	0.04	[0.26, 0.43]	[0.23, 0.46]

- Analyst information is most valuable in the **first week following earnings announcement.**

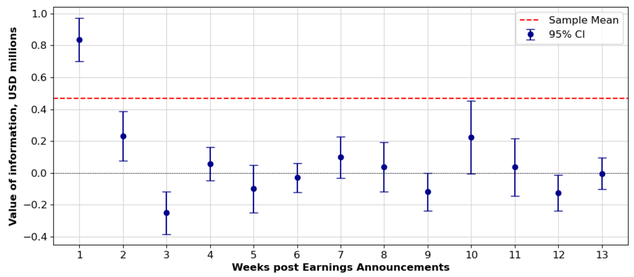


Fig 3. Analyst information value following earnings announcements

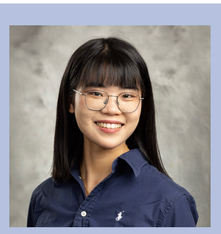
- Analyst reports add **significant incremental value** beyond EA.

	Transcripts	Reports	Reports + Transcripts	Diff
	(1)	(2)	(3)	(3) - (1)
R^2_{OOS}	4.20%	9.72%	11.96%	7.76%
t-stat	5.16	3.24	6.42	5.12

Takeaways

- **Textual information** in analyst reports explains 10.19% of contemporaneous stock returns out-of-sample, a value that is **economically more significant than quantitative forecasts**.
- Analysts' **income statement analyses** account for more than half of the reports' explanatory power.
- Early acquisition of analysts' reports yields **significant profits**.
- Analysts' information value **peeks following earnings announcements**, highlighting their **vital role in interpreting new financial data**.

Listen to my paper podcast here!



Contact

Linying Lv
llyu@wustl.edu
https://www.linkedin.com/in/linying-lyu-b70073289/